



**Fakultät für  
Informatik und Mathematik**

# **Bachelorarbeit**

über das Thema

**Entwurf und Implementierung eines Chrome Plugins zur  
automatischen Anreicherung von Webseiten mit kulturellen  
Inhalten**

**Autor:** Mathias Möller  
moellerm@fim.uni-passau.de

**Prüfer:** Prof. Granitzer

**Abgabedatum:** 24.08.2015

## **I Kurzfassung**

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## **Abstract**

Das ganze auf Englisch.

## II Inhaltsverzeichnis

<b>I</b>	<b>Kurzfassung</b>	<b>I</b>
<b>II</b>	<b>Inhaltsverzeichnis</b>	<b>II</b>
<b>III</b>	<b>Abbildungsverzeichnis</b>	<b>IV</b>
<b>IV</b>	<b>Tabellenverzeichnis</b>	<b>V</b>
<b>V</b>	<b>Abkürzungsverzeichnis</b>	<b>V</b>
<b>1</b>	<b>Einleitung und Motivation</b>	<b>1</b>
<b>2</b>	<b>Related Work</b>	<b>3</b>
2.1	Unterschiede und Gemeinsamkeiten zu JITIR-Agents . . . . .	3
2.1.1	Proaktiv vs. User Interaction . . . . .	4
2.1.2	Informationen darstellen in „Nonintrusive Manner“ . . . . .	4
2.1.3	Awareness of user's local context . . . . .	4
2.2	Vergleich mit existierenden JITIR Agents . . . . .	4
2.2.1	Remembrance Agent (in EMACS Editor) . . . . .	4
2.2.2	Margin? Web Plugin ähnlich wie EEXCESS . . . . .	4
2.2.3	Watson . . . . .	4
2.2.4	EEXCESS? . . . . .	4
2.3	Text Retrieval Algorithmen . . . . .	4
2.4	Unterschiede und Gemeinsamkeiten zu „Automatic help systems“ (z.B. Microsoft Office Assistant - i Domain spezifisch) . . . . .	4
<b>3</b>	<b>Konzept</b>	<b>5</b>
3.1	Warum kein Proactiver JITIR-Agent? . . . . .	5
3.2	Anzeige der Ergebnisse . . . . .	5
3.3	Erklärung der Such-Anfragen Generierung . . . . .	5
3.4	Anpassen der Suchanfrage durch den Nutzer . . . . .	5
3.5	Verbesserung der Suchanfrage z.B. durch maschinelles Lernen . . . . .	5
<b>4</b>	<b>Implementierung</b>	<b>5</b>
4.1	Verwendung von AngularJS für alle Komponenten des Plugins . . . . .	5
4.2	Bau der GUIs . . . . .	5
4.3	Einbindung der REST-Services . . . . .	5
<b>5</b>	<b>Evaluierung?</b>	<b>5</b>
<b>6</b>	<b>Future Work</b>	<b>5</b>
6.1	Alternative Algorithmen zur Textanalyse . . . . .	5
6.2	Implementierung einer automatischen Suchanfragen-Verbesserung durch maschinelles Lernen . . . . .	5
6.3	Verbesserung der Ergebniss-Güte . . . . .	6
6.4	Anpassung der Anwendung auf mobile Nutzung . . . . .	6

<b>7 Conclusion</b>	<b>6</b>
<b>8 Quellenverzeichnis</b>	<b>7</b>
<b>Anhang</b>	<b>I</b>

### **III Abbildungsverzeichnis**

## **IV Tabellenverzeichnis**

## **V Abkürzungsverzeichnis**

# 1 Einleitung und Motivation

Suchmaschinen gehören heutzutage zu den meistbesuchten Seiten im Internet. Google<sup>1</sup> ist mit 2.000.000.000.000 Suchanfragen pro Jahr (TODO: Quelle) nicht nur die beliebteste Suchmaschine, sondern auch die meist besuchte Seite im Internet. Auch die chinesische Suchmaschine Baidu<sup>2</sup> ist in den oberen fünf Plätzen vertreten [AI].

- Auf XX% der Webseiten wird über eine Suchmaschine zugegriffen
- Menschen holen Informationen über Suchmaschinen, nicht mehr aus Bibliotheken oder Lexika
- Woher Quellen?

Um so mehr sie an Bedeutung gewinnen, umso größer ist der Bedarf an bessern Suchmaschinen [Law00].

Doch der Funktionsumfang klassischer Suchmaschinen ist begrenzt. Sie können den Kontext des Benutzers in ihren Queries? nicht miteinbeziehen. Weiterhin verlangen sie vom Nutzer, seine gegenwärtige Arbeit einzustellen, die Webseite der Suchmaschine aufzurufen und eine Suchanfrage zu formulieren. So wird die Konzentration auf den eigentlichen Arbeitsschritt gestört. Im Schnitt sind nur 67% der Suchanfragen an Google erfolgreich (TODO: Quelle). Das heißt in circa einem Drittel der Fälle bleibt der Aufruf der Suchmaschine ohne die erwarteten Erfolge. Zu der verlorenen Zeit addiert sich auch noch die Zeit, die der Nutzer braucht um sich seine Arbeit wieder ins Kurzzeitgedächtnis zu holen.

Um diesen Problemen entgegen zu wirken wurden neue Wege zur Informationsgewinnung entwickelt. Eine Möglichkeit sind so genannte Just-in-time Information Retrieval Agents (JITIR-Agents) [RM00]. Sie beobachten im Hintergrund den Kontext des Benutzers und versuchen aus den so erhaltenen Informationen eine Suchanfrage an eine Datenbank oder ein Recommender-System zu schicken. Die gewonnenen Informationen werden dann möglichst unaufdringlich dem Benutzer angezeigt. Er kann sich nun entscheiden diese Informationen genauer zu betrachten oder mit seiner Arbeit fortzufahren. Die kognitive Belastung bleibt hierbei sehr gering. JITIR-Agents reduzieren auf diese Weise enorm den Aufwand Informationen zu finden [RM00].

Durch ihre Funktionsweise sind sie jedoch nicht so exakt wie klassische Suchmaschinen, da sie nur 'erraten' können, was den Benutzer gerade interessiert. Wenn ein Benutzer einer genauen Vorstellung von der Suchanfrage oder den Ergebnissen hat, haben klassische Suchmaschinen Vorteile gegenüber den JITIR-Agents [RM00].

---

<sup>1</sup><https://www.google.com/>

<sup>2</sup><http://www.baidu.com/>

Ein weiteres Problem des Internets ist es, dass die Standards, die die simple und leicht skalierbare Architektur des Netzwerks ermöglichen, den Einsatz von fortgeschritteneren Hypermedia-Technologien verhindern [Bou99]. Eine Möglichkeit, das Internet um Funktionen zu erweitern ist Web Augmentation (WA). Diaz [Día12] beschreibt WA als den Versuch, statt eine neue Technologie zu entwickeln, neue Funktionalität auf eine gerenderte Webseite zu setzen:

In some sense, WA is to the Web what Augmented Reality is to the physical world: layering relevant content/layout/navigation over the existing Web to customize the user experience. [Día12]

Solche WA-Tools interagieren mit einem Web Server, Http-Proxy oder dem Browser der Nutzers, um so Inhalte oder Navigationselemente direkt in die angezeigte Webseite einzufügen. Auf diese Weise erlauben sie es, die limitierten Möglichkeiten des World Wide Webs zu Gunsten des Nutzers anzureichern (Quelle: Kennth anderson oder Bouvin).

Ein Beispiel für ein solches WA-Tool sowie für ein JITIR System ist das EEXCESS Chrome Plugin<sup>3</sup>. Es analysiert die aktuell besuchten Seiten des Nutzers und schlägt ihm auf Grund der erlangten Daten am Rand des Browser-Fensters weiterführende Quellen aus der Europeana-Datenbank<sup>4</sup> vor.

Ziel dieser Bachelorarbeit ist es, auf Basis des EEXCESS Plugins ein Chrome Plugin zu entwerfen und zu implementieren, welche eine Webseite in einzelne Paragraphen aufteilt und zu jedem dieser Paragraphen kulturelle Inhalte der Europeana-Datenbank vorschlägt. Dabei soll das Design des Plugins helfen, die bei der in Kapitel 5 erläuterten Evaluierung des EEXCESS Plugins aufgetauchten Probleme zu beheben und Schwachstellen zu verbessern.

Im zweiten Teil dieser Arbeit wird auf vergleichbare Technologien eingegangen und ihre Gemeinsamkeiten und Unterschiede zu diesem Projekt. Anschließend wird das Konzept und die Implementierung des Projekts beschrieben und das Ergebnis evaluiert. Der sechste Teil der Ausarbeitung widmet sich Möglichkeiten für zukünftige Projekte und Verbesserungen. Zuletzt wird ein Fazit gezogen und die Ergebnisse der Arbeit analysiert.

---

<sup>3</sup><http://eexcess.eu/results/chrome-extension/>

<sup>4</sup><http://www.europeana.eu/portal/>



## 2 Related Work

Rhodes[RM00] definiert JITIR-Agents als eine Klasse von Programmen, die dem Benutzer weiterführende Informationen basierend auf seinem lokalen Kontext anzeigen. Dabei beschränkt sich der überwachte Kontext meist auf die virtuelle Umgebung des Benutzers, wie E-Mail, Webseiten und geöffnete Dokumente. Als Kerneigenschaften von JITIR-Agents nennt Rhodes Selbstständigkeit, die Fähigkeit Informationen in einer leicht zugänglichen und gleichzeitig unaufdringlichen Weise darzustellen und das Bewusstsein über den Kontext des Benutzers[RM00]. Das im Zuge dieser Bachelorarbeit entworfene Programm (im Folgenden als Jarvis bezeichnet) erfordert vom Nutzer die explizite Aufforderung um nach Informationen zu suchen. Gründe für diese Entscheidung werden im dritten Kapitel näher betrachtet. Trotz dieses Widerspruchs zu Rhodes Definition lässt sich Jarvis am besten mit dieser Klasse von Programmen vergleichen.

### 2.1 Unterschiede und Gemeinsamkeiten zu JITIR-Agents

studies in computer response time indicate that small increases in the effort required to perform a task can have large effects on whether a person will bother acting at all. miller argues that for many tasks more than two seconds of response delay is unacceptable and will result in fewer uses of a particular tool, even at the cost of decreased accuracy - i robert miller same short-term memory limitations miller discusses also apply to performing subtasks that distract from a primary task. for example when performing a search for information about a digressing, a person needs to use short-term memory to keep his or her place in the larger framework of the task. the amount of short-term memory required, and thus the amount of effort required in the task, will depend on a number of factors including the complexity of the digression, the amount of time required to complete the task, and the similarity of the digression to the primary task. - i applied to the information search domain, these theories suggest that if the cost of finding and using information is more than the expected value of that information, then the search will not be performed. Jitir agents greatly reduce the costs of searching for information by doing most of the work automatically. the downside is that queries are automatically generated and therefore will not be as exact as would a human-generated query.

will not bother perform a search because a lack of time. information he has is already "good enough. person expects a search will not turn up anything useful. rhodes

### **2.1.1 Proaktiv vs. User Interaction**

### **2.1.2 Informationen darstellen in „Nonintrusive Manner“**

### **2.1.3 Awareness of user's local context**

## **2.2 Vergleich mit existierenden JITIR Agents**

### **2.2.1 Remembrance Agent (in EMACS Editor)**

Zeigt Quellen an auf Basis des geschriebenen Texts, Benutzer kann Suchanfrage dann auch noch manuell anpassen/verfeinern - ı Vorteil von Suchmaschinen wird mit integriert

### **2.2.2 Margin? Web Plugin ähnlich wie EEXCESS**

### **2.2.3 Watson**

### **2.2.4 EEXCESS?**

## **2.3 Text Retrieval Algorithmen**

Term Frequency/Inverse Document Frequency algorithm Text rank

## **2.4 Unterschiede und Gemeinsamkeiten zu „Automatic help systems“ (z.B. Microsoft Office Assistant - ı Domain spezifisch)**

Domain spezifisch vs. Domain unabhängig

## 3 Konzept

### 3.1 Warum kein Proactiver JITIR-Agent?

- ¿ API Limitierung und decrease cognitive load - ¿ Benutzer entscheidet ob er weitere Informationen erhalten möchte - ¿ Benutzer kann Suchanfrage erst anpassen (Nachteil von Margin Notes (Paper 4))

### 3.2 Anzeige der Ergebnisse

### 3.3 Erklärung der Such-Anfragen Generierung

### 3.4 Anpassen der Suchanfrage durch den Nutzer

### 3.5 Verbesserung der Suchanfrage z.B. durch maschinelles Lernen

## 4 Implementierung

### 4.1 Verwendung von AngularJS für alle Komponenten des Plugins

### 4.2 Bau der GUIs

- ¿ Darf den Benutzer nicht zu sehr ablenken - ¿ Ergebnisse müssen in der Nähe ihrer „Quelle“ angezeigt werden (proximity compatibility principle) - ¿ Benutzer muss klar zwischen Webseite und Augmentation unterscheiden können - ¿ bunt, auffälliges Design - ¿ Ramping interface: Mehr Benutzerinteraktionen führen zu mehr angezeigten Informationen (Erklärung der Stages)

### 4.3 Einbindung der REST-Services

## 5 Evaluierung?

Aufgaben die Benutzer mit EEXCESS Lösen mussten müssen sie jetzt mit Redesign lösen. Vergleich der Ergebnisse?

## 6 Future Work

### 6.1 Alternative Algorithmen zur Textanalyse

### 6.2 Implementierung einer automatischen

#### Suchanfragen-Verbesserung durch maschinelles Lernen

- Mehr kontextuelle Informationen miteinbeziehen
- Such-Profil des Nutzers erstellen

### **6.3 Verbesserung der Ergebniss-Güte**

- durch Query Expansion
- durch Filtern der Ergebnisse (mehr Präzision da Ausbeute bei JITIR nicht so relevant)

### **6.4 Anpassung der Anwendung auf mobile Nutzung**

## **7 Conclusion**

- Steigerung der Effektivität und Produktivität von wissenschaftlichem Arbeiten
- Starke Effektivitätssteigerung durch Punkte aus Future Work möglich?

## 8 Quellenverzeichnis

- [AI] ALEXA INTERNET, Inc. 1996 2.: *Alexa*. <http://www.alexa.com/topsites>. – Zugriff: 13.05.2015, Archiviert mit WebCite®: <http://www.webcitation.org/5hgZUZacN>
- [Bou99] BOUVIN, Niels O.: Unifying strategies for Web augmentation. In: *Proceedings of the tenth ACM Conference on Hypertext and hypermedia: returning to our diverse roots: returning to our diverse roots* ACM, 1999, S. 91–100
- [Día12] DÍAZ, Oscar: Understanding web augmentation. In: *Current Trends in Web Engineering*. Springer, 2012, S. 79–80
- [Law00] LAWRENCE, Steve: Context in web search. In: *IEEE Data Eng. Bull.* 23 (2000), Nr. 3, S. 25–32
- [RM00] RHODES, Bradley J. ; MAES, Pattie: Just-in-time information retrieval agents. In: *IBM Systems journal* 39 (2000), Nr. 3.4, S. 685–704

## Anhang

### GUI Screenshots

Unterkategorie, die nicht im Inhaltsverzeichnis auftaucht.

# Erklärung

Hiermit versichere ich, dass ich meine Abschlussarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Datum:

.....

(Unterschrift)